



**Título: NACluster: Resolvendo Entidades em Larga Escala a partir de Múltiplos Catálogos de Astronomia.**

**Data: 30/08/2016 Horário: 09h Local: Sala de Seminários do Bloco 952 - Campus do Pici**

Resumo:

Levantamentos astronômicos usam instrumentos poderosos para navegar o céu e identificar objetos de interesse dentro da região pesquisada. Objetos do céu são caracterizados individualmente com coordenadas espaciais, identificando a sua posição no céu, além de outros atributos descritivos. Gerar uma visão integrada do céu com base em catálogos produzidos por diferentes levantamentos enfrenta um difícil problema de casamento de catálogos. Devido às variações na calibração dos instrumentos de captura, a posição de um único objeto celeste pode variar de um catálogo para outro. Além disso, em determinadas regiões densas do céu, este problema é agravado por um grande número de potenciais candidatos ao casamento para cada objeto. As abordagens tradicionais para lidar com esse problema usam uma distância limite de epsilon para reduzir o número de candidatos correspondentes. Além disso, elas adotam uma abordagem para casar 'n' catálogos aos pares, inferindo transitividade no casamento, o que nem sempre possuem. Nesta tese, apresentamos o NACluster, um algoritmo de clusterização não-supervisionado para realizar o casamento de vários catálogos. A estratégia de casamento do NACluster estende o tradicional algoritmo de clusterização k-means relaxando o número k de clusters (ou seja, objetos celestes casados). Também propomos o ParallelNACluster, uma versão paralela do NACluster que se aproveita do particionamento dos dados de entrada, e aceita grandes volumes de dados mesmo utilizando um conjunto de hardware de pequeno porte. Além disso, propomos o SCIBoundary, uma nova estratégia para tratamento do casamento de objetos espacialmente separados em partições de dados vizinhas. O SCIBoundary permite que obtenhamos

resultados equivalentes entre o NACluster e o ParallelNACluster. Nesta tese, também apresentamos o AODP, um workflow para realizar o particionamento dos dados em disco local e executar o casamento dos mesmos em ambiente distribuído através do ParallelNACluster. Nossos experimentos mostram a eficiência das estratégias centralizada e paralela através de excelentes índices de acertos obtidos nas clusterizações, bem como a eficiência do SCIBoundary no tratamento das fronteiras.

Banca:

- Prof. Dr. José Antônio Fernandes de Macedo (MDCC/UFC)
- Prof. Dr. Fábio André Machado Porto(LNCC/RJ)
- Prof. Dr. Marco Antônio Casanova(PUC/RJ)
- Prof. Dr. Ricardo Lourenço Correia Ogando(ON)
- Profa. Dra. Vânia Maria Ponte Vidal(MDCC/UFC)
- Prof. Dr. José Maria da Silva Monteiro Filho (MDCC/UFC)